

# **The existential/canonical alternation in Brazilian Portuguese**

**Results from a corpus study**

Alex Bratkievich  
Linguistics H195  
Spring 2006  
Professor Line Mikkelsen

## 0. Introduction

In modern linguistics, existential constructions (e.g. *There's a cat on the mat*) have been the object of intense research since at least Milsark (1974). Their alternation with canonical constructions (e.g. *A cat is on the mat*) is peculiar because the two are semantically equivalent, in the restricted sense of having the same truth-conditions. If languages give the option between two constructions that have the same meaning, the preference for one or the other must have some communicative relevance. It is still debated whether the factors determining the choice of existentials over canonicals in certain situations are the same cross-linguistically. A strong indication that this might be the case is the fact that existential sentences across languages share certain properties, such as the DEFINITENESS EFFECT. This is the name given to the strong preference of existentials to have an indefinite DP<sup>1</sup> in the post-verbal position (usually referred in the literature as the PIVOT).

This thesis focuses on the existential/canonical alternation in Brazilian Portuguese (BP), using corpus data to establish the frequency and distribution of DPs in the two constructions. The remainder of this chapter will describe the facts relevant to canonical and existential constructions in BP. Chapter 1 presents three standard analyses proposed for existentials: as a sub-class of unaccusatives (Perlmutter 1976, Zubizarreta 1982); as impersonal constructions (Franchi et al. 1998, Viotti 1999); and as locatives (Freeze 1992, Kempchinsky 1995). Chapter 2 deals with previous accounts given for the definiteness effect (Milsark 1977, Mikkelsen 2002). Chapter 3 introduces the methodology used for the corpus search, while Chapter 4 presents the results, which show the non-categorical nature of the definiteness effect and suggest that previous explanations cannot fully account for the entire distribution of DPs. Finally, chapter 5 shows further evidence from the corpus data in favor of an account within the framework of Optimality Theory.

### 0.1 Canonical and existential constructions in BP.

BP is usually assumed to have a basic SVO order, although VSO sentences are quite common. For the purposes of this thesis, the CANONICAL construction is the copular construction

---

<sup>1</sup> Even though I remain agnostic as to the adequacy of the DP hypothesis, for the sake of convenience I've used DP throughout this thesis. However, I don't believe that the DP/NP distinction is relevant to the issues discussed here.

with *estar*, which is exemplified in (1). The order is SVX, where X stands for any copular complement, a Locative PP in (1). The verb *estar* agrees in number with the subject *um gato*.

- (1) Um gato está            no    tapete.  
a    cat    be.PRES.3SG in.the mat  
"A cat is on the mat."

The canonical construction alternates with the existential construction, which is exemplified in (2) and (3). Unlike in English, the existential verbs don't agree with the pivot; *ter* and *haver* appear with default 3SG agreement.

- (2) Há                    um gato no    tapete.  
have.PRES.3SG a    cat    in.the mat  
"There's a cat on the mat."
- (3) Tem                    dois gatos no    tapete.  
have.PRES.3SG two    cats    in.the mat  
"There are two cats on the mat."

The issue of defining what constitutes an existential sentence is not trivial and it will be discussed at length in Chapter 1. However, for the purposes of this thesis, I will follow Viotti (1999) in restricting the term EXISTENTIALS to impersonal constructions with the verbs *ter* and *haver*.

*Ter* and *haver*, which are both translated as "have," evolved from two different Latin verbs: *habere* and *tenere*. According to Franchi et al. (1998), both verbs entered the Portuguese language (as *aver* and *teer*) with a possessive sense. However, there are strong indications that *aver* was preferred to express inalienable possession whereas *teer* was preferred for circumstantial possession. Between the XIV and XVI centuries, *teer* became the preferred verb for all possessive predications. As for existential constructions, until the XIV century, the preferred verb was *aver* (which alternated with *seer*, "be"). The first attested uses of *teer* in existential constructions occur in the XVI century, only after it has become the predominant verb in possessive constructions.

Nowadays, the preferred verb for existentials in spoken BP is *ter*, which appeared in 63% of the existential sentences found in a sample of the Projeto NURC-RJ (Rio de Janeiro) spoken corpus from the 70s; in a sample from the 90s, the percentage of use of *ter* had spiked to 76%

(Duarte 2003, citing Callou e Avelar 2001). Abstracting away from sociolinguistic factors (education level, age, etc.), both verbs can, in theory, be used interchangeably<sup>2</sup>.

According to Franchi et al. (1998), *haver* has been completely emptied of its lexical meaning and has specialized as the functional head of existential constructions, whereas *ter* has retained its lexical content in specific contexts<sup>3</sup>. *Ter*, for example, is used to express both inalienable (as in (4)) and circumstantial possession (as in (5)<sup>4</sup>):

- (4) Cryolophosaurus ellioti tinha uma crista na testa.  
Cryolophosaurus ellioti have.PAST.3SG a.FEM crest in.the.FEM forehead.  
"Cryolophosaurus ellioti had a crest in the forehead."
- (5) Camila tem uma casa em Campos.  
Camila have.PRES.3SG a.FEM house in Campos.  
"Camila has a house in Campos."

In Chapter 4, I will present the results of a quantitative study performed on a written corpus, which reveal a different frequency of use for *ter* and *haver* from the spoken BP data.

## 1. Standard analyses for BP existentials

Several analyses have been proposed for BP existentials; each of them considers existentials as part of a different natural class (the unaccusative constructions, the impersonal constructions, the locative paradigm). This chapter will briefly consider these three proposals.

### 1.1 Existentials as unaccusatives

Several authors (Thomas 1969, Perlmutter 1976, Zubizarreta 1982, Nascimento 1984, Silva 1994<sup>5</sup>) have analyzed existentials as belonging to the same natural class as UNACCUSATIVE constructions (i.e. intransitive verbs whose only argument shares thematic properties with the

---

<sup>2</sup> This assumption will be qualified in Section 5.3.

<sup>3</sup> Both *haver* and *ter* are also used as perfective auxiliaries, but that use is not directly relevant for the present study of their main verb uses.

<sup>4</sup> Examples #4 and 5 were taken from the CETENFolha Corpus.

<sup>5</sup> Unfortunately, I didn't have access to Nascimento (1984) and Silva (1994), which are PhD dissertations from Université de Paris and Université de Genève respectively. For this section, I am following Franchi et al. (1998) in their presentation of Nascimento's arguments to consider existentials and unaccusatives as a natural class (although they reach a different conclusion).

object of transitive verbs). This proposal is based on semantic, syntactic and pragmatic considerations.

Perlmutter (1976) –citing and expanding the original list of 7 verbs in Thomas (1969)<sup>6</sup> to a total of 13 verbs– includes among the class of Portuguese existential sentences<sup>7</sup> constructions with verbs such as *acontecer* ("happen"), *aparecer* ("appear"), *chegar* ("arrive"), *existir* ("exist"), *faltar* ("lack"), *ocorrer* ("occur"), *sobrar* ("be left"), *sumir* ("disappear"), *surgir* ("arise") and *vir* ("come"). He supports this claim with two observations:

- a. "most of these verbs express either an affirmation or denial of existence, including coming into existence and ceasing to exist. The notion of existence involved may be existence with respect to someone's frame of reference..."
- b. "Most of the Portuguese sentences cited above can be expressed in English by means of sentences with existential there" (p. 95)<sup>8</sup>.

Not only do these verbs appear in similar contexts as existentials; they can sometimes be used interchangeably with only slight nuances in the difference between the meanings. The clearest case are sentences with *existir*:

(6) Cem anos atrás não **tinha** [essa] histeria<sup>9</sup>  
hundred years ago NEG **have.PAST.3SG** this.FEM hysteria  
"A hundred years ago, there wasn't this hysteria."

(7) Cem anos atrás não **existia** essa histeria  
hundred years ago NEG **exist.PAST.3SG** this.FEM hysteria  
"A hundred years ago, this hysteria didn't exist."

This interchangeability is reflected when speakers hesitate about which form to use, as can be seen in (8) and (9):

---

<sup>6</sup> Thomas list of 7 verbs was: *existir* ("exist"), *faltar* ("lack"), *sobrar* ("be left"), *ficar* ("remain"), *aparecer* ("appear"), *surgir* ("arise"), *sumir* ("disappear").

<sup>7</sup> Thomas' work was based on BP data. Perlmutter's work confirms the observations presented in Thomas (1969) for European Portuguese (EP) and expands the list of verbs to 13. Even though Perlmutter's work is about EP, it seems that its validity for BP has been assumed as correct (or taken for granted) in the literature.

<sup>8</sup> The article itself is concerned with the subject status of the post-verbal DP and, after this brief observation in his introduction, Perlmutter doesn't mention existential constructions with *ter* or *haver* again.

<sup>9</sup> Examples #25a. and b. in Franchi et al. (1998); the original sentence (with *ter*) is from the spoken corpus Projeto NURC-SP (São Paulo). (7) is a modification of example #25b.



	Total	Non-agreement examples		Agreement examples	
		Number	Percentage	Number	Percentage
Unaccusatives with post-posed subject	30	7	23.33%	23	76.66%
Existentials with <i>ter/haver</i> <sup>14</sup>	40	38	95%	2	5%

**TABLE 1: Agreement in unaccusative and existential constructions.**

Other syntactic similarity between unaccusatives with post-posed subjects and existentials is that there is usually a "secondary"<sup>15</sup> predication associated with the post-verbal DP. In sentence (9), this predication appears in the form of a locative PP (*diante de mim*); in sentence (10), as a restrictive relative clause (*que não são explorados*). The relationship between the post-posed subject and the PP or RRC seems to be parallel to the relationship between the pivot and the CODA, which is the constituent that follows the pivot in an existential, such as *no tapete* in (2) and (3).

Another common characteristic is what Franchi et al. (1998) call a SPATIAL-TEMPORAL ANCHORAGE ("*uma ancoragem espaço-temporal*"). In the examples above, this anchorage is sometimes realized in the coda (*no tapete* in (2) and (3), *no hospital* in (8), *diante de mim* in (9)), as a pre-posed locative or temporal phrase (*cem anos atrás* in (6) and (7)), or it might be left implicit (as in (10)).

Finally, at least some unaccusative constructions evidence the definiteness effect, i.e. there is a preference for these post-posed subjects to be indefinite, as Table 2<sup>16</sup> indicates. However, it must be noted that the evidence suggests there might be a definiteness effect in the case of constructions with *existir*, whereas the number of indefinite and definite DPs is the same for other unaccusative constructions with post-posed subjects (the column "Others" in Table 2).

<sup>14</sup> It should be noted that, although there were 490 existentials in the corpus, only 40 of them are suitable to check if the verb agrees or not with the pivot. This is due to the homophony between the 3SG PRESENT INDICATIVE form of *ter* (*tem*) and the 3PL PRESENT INDICATIVE form (*têm*), which are differentiated only orthographically.

<sup>15</sup> As will be seen in Section 1.2, the relationship between pivot and coda is, in fact, the main predication.

<sup>16</sup> *Tabela A* from Franchi et al. (1998), page 27.

		<i>Existir</i>	Others	Sub-total	Total	Percentages
Indefinite	Indefinites	53	22	75	101	60.11%
	Bare-NP Indefinites	24	2	26		
Definite	Bare-NP Definites	11	2	13	67	39.88%
	Definites	31	23	54		
Instances		119	49	168		

**TABLE 2: Indefinite and definite DPs in the post-verbal position of unaccusatives.**

It has also been noted that the post-verbal position is a position of presentational focus. Guéron (1980) uses this fact to distinguish PREDICATIONS, in which the VP is in the c-command domain of the subject, from PRESENTATION S(entence)s, in which the subject is in the scope of the verb in LF. Predications presuppose the existence of the entity referred by the subject. Presentation Ss introduce a new discourse referent. Guéron noted that in some constructions, such as the English existential and the Italian NP Postposition, the presentation S configuration is already evident in surface structure. Under these assumptions, if we extend Guéron's observations to other languages, there is another reason for grouping existentials and unaccusatives with post-posed subjects: both constructions introduce a new discourse referent, the post-verbal DP.

### 1.2 Existentials as impersonals

It seems that existential sentences in the semantic sense (i.e. sentences which are used to express the existence of the entity referred by the DP) cannot be grouped under one single syntactic construction. There are existentials with *ter* and *haver* on one side, and existentials with *existir* on the other. Moreover, If we assume that the notion of existence could be the existence with respect to someone's frame of reference (as Thomas and Perlmutter do) or the existence as a discourse referent (as Guéron does), unaccusative constructions with other verbs (not only with *existir*) can be argued to be "existentials." The previous section explored several of the similarities between these constructions. This section will focus on the reasons that lead Franchi

et al. (1998) and Viotti (1999) to include existentials among the class of the impersonal constructions.

These authors criticize the rigid dichotomy proposed by Guéron (1980) based on alternations such as the following, in which it is very difficult to identify any meaning differences between the presentation S configuration (post-verbal subject, as in (12)) and the predication configuration (pre-verbal subject, as in (11)).

(11) Eu acho que não existe esse valor<sup>17</sup>

I find that NEG exist.PRES.3SG this value

"I think this value doesn't exist" (talking about the social value of the humanities).

(12) Eu acho que esse valor não existe

I find that this value NEG exist.PRES.3SG

"I think this value doesn't exist."

Moreover, because *ter* and *haver* have been emptied of any lexical content in their existential use, the only predication found in an existential is between the pivot and the coda. On the other hand, in most examples of unaccusative verbs with post-posed subjects, the verb still retains most or all of its lexical content. Therefore, there are two predicative relationships in the sentences with these verbs: one between the subject and the verb, another between a secondary predication and the subject.

Franchi et al. (1998) argue that existentials are closer syntactically to IMPERSONAL CONSTRUCTIONS (i.e. constructions without an external argument). In BP, it is possible for some verbs to experience a process of "impersonalization," as illustrated by (13) and (15). (13) shows that this is possible with transitive verbs such as *dar*, "give" (cf. its use in (14)); (15) shows that unaccusative verbs can also appear in impersonal constructions (cf. the use of *chegar*, "arrive," in (16)).

(13) Imagina se dá um aperto, ou acontece qualquer coisa...<sup>18</sup>

imagine.IMP if give.PRES.3SG a tight.spot or happen.PRES.3SG any thing

"Imagine if a tight spot occurs, or anything happens" (talking about the need for savings)

Literally: "Imagine if (it) gives a tight spot, ..."

<sup>17</sup> Examples #44a. and b. in Franchi et al. (1998); the original sentence (with post-posed subject) is from the spoken corpus Projeto NURC-PA.

<sup>18</sup> Example #54 in Franchi et al. (1998), taken from Projeto NURC-PA.

(14) Essas datas me dão um aperto no coração...<sup>19</sup>  
 that.PL date.PL 1SG.DAT give.PRES.3PL a tight.spot in.the heart  
 "Those dates give me a heartache."

(15) Quando chega na hora de comprar mais roupa, ela não  
 when arrive.PRES.3SG in.the.FEM time of buy.INF more clothing she NEG  
 pagou ainda aquela<sup>20</sup>  
 pay.PAST.3SG yet that.DISTAL  
 "When the time to buy more clothing arrives, she hasn't paid for that one yet."  
 Literally: "When (it) arrives in the time to buy more clothing, ..."

(16) Depois do meio-dia, chega a hora do garrafão.<sup>21</sup>  
 after of.the midday arrive.PRES.3SG the.FEM time of.the flagon  
 "After noon, the time of the flagon arrives."

The internal argument of the impersonal constructions is usually the focus, which means that impersonals, unaccusatives, and existentials have a presentational function. Franchi et al. (1998) also notice that there is a preference for indefinite DPs in the post-verbal position of the three constructions. However, the key difference that sets apart impersonals and existentials from unaccusatives is the lack of agreement between the verb and the post-verbal DP<sup>22</sup>.

Another syntactic consideration is that the post-verbal DP of existentials and impersonals gets Accusative Case, whereas the post-verbal DP of unaccusatives gets Nominative Case<sup>23</sup>. Example (17) shows that the resumptive pronoun in place of the left-dislocated pivot appears in the accusative form.

<sup>19</sup> Example taken from <http://br.search.yahoo.com>.

<sup>20</sup> Example #68 in Franchi et al. (1998), taken from Projeto NURC-PE (Recife).

<sup>21</sup> Example taken from the CETENFolha Corpus.

<sup>22</sup> It should still be noticed that there is some degree of variation and both unaccusative constructions in which the verb does not agree with their subject as well as existentials in which the verb agrees with the pivot can be found in the language. The direction of the change and the reasons for this variation are, unfortunately, outside the scope of this thesis.

<sup>23</sup> It is very difficult to find naturally-occurring data for this distinction because morphological case is only evident in the clitic system. However, the 3SG clitics are strongly dispreferred in BP (the Nominative/Accusative distinction is disappearing and stressed pronouns –in their Nominative form– are used even when they stand for an object).

(17) Cerimônias religiosas,       houve-as                               e   muito concorridas<sup>24</sup>  
ceremonies religious.FEM.PL have.PAST.3SG-3PL.FEM.ACC and very popular.FEM.PL  
"Religious ceremonies there were, and very popular."

Literally: "Religious ceremonies, there were them and very popular."

Consequently, even though constructions with *existir* and other verbs might be semantically and pragmatically related to the constructions with *ter* and *haver*, in this thesis I only consider the latter.

### 1.3 Existentials as locatives

Freeze (1992) proposes a third alternative: existentials are members of the LOCATIVE PARADIGM, which includes also PREDICATE LOCATIVES and *HAVE* PREDICATIONS. These three constructions are derived from "a single and maximally simple abstract syntactic structure" (p. 554). This accounts straightforwardly for the structural similarity of BP existentials and *have* predications (as can be seen from (2) through (5) above, repeated below as (18) through (21)). The main difference is that existentials lack an external argument. In fact, if the subject were not overtly expressed in (21), out of context it would not be possible to distinguish between the two possible readings ("*There's a house in Campos*" and "[*s/he*] *has a house in Campos*").

(18) Há                               um gato no       tapete.  
have.PRES.3SG a cat in.the mat  
"There's a cat on the mat."

(19) Tem                               dois gatos no       tapete.  
have.PRES.3SG two cats in.the mat  
"There are two cats on the mat."

(20) Cryolophosaurus ellioti tinha                               uma crista na                               testa.  
Cryolophosaurus ellioti have.PAST.3SG a.FEM crest in.the.FEM forehead.  
"Cryolophosaurus ellioti had a crest in the forehead."

---

<sup>24</sup> Example #82 from Franchi et al. (1998).

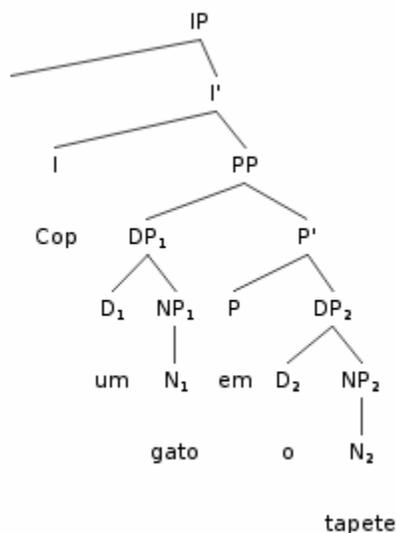
- (21) Camila tem uma casa em Campos.  
 Camila have.PRES.3SG a.FEM house in Campos.  
 "Camila has a house in Campos."

The third component of the locative paradigm, predicate locatives (referred to as "canonicals" in this thesis), contrast with the other two in using a different verb, *estar*, as exemplified in (1) above, repeated below as (22):

- (22) Um gato está no tapete.  
 a cat be.PRES.3SG in.the mat  
 "A cat is on the mat."

According to Freeze's analysis, (22) is derived from a structure like (23):

- (23)



The copula is base-generated in  $I^\circ$  as a bundle of features which is spelled out at PF as *estar*.  $P^\circ$  is the head of a predicate phrase that selects two arguments: a theme (*um gato*), which is base-generated in Spec-PP, and a location (*o tapete*), which is base-generated as the sister of  $P^\circ$ . The theme moves to Spec-IP and the result is the predicate locative.

The subject of the *have* predication is a syntactic location. In the derivation, the locative argument moves to subject position. The absence of a preposition is explained as the result of  $P^\circ$  incorporating into  $I^\circ$ , which as a result is spelled out as *ter*.

The existential is derived without any movement, with the theme and the location appearing in their base-generated position.

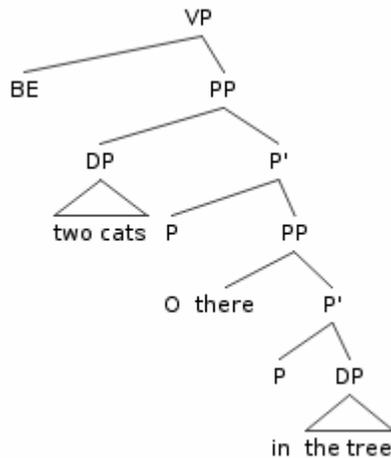
### 1.3.1 Deriving the existential

As Viotti (1999) points out, if *ter* is the spell-out of the copula plus an incorporated P°, how is the existential verb spelled out as *ter*? As can be seen in (23), the proposed underlying structure for (24), there's no movement involved; the locative DP is in its base-generated position. What P° is incorporated into the copula? It can't be the P° of the predicative phrase because that P° remains in situ.

- (24) Tem                    um gato no     tapete.  
 have.PRES.3SG a    cat    in.the mat  
 "There's a cat on the mat."

It seems that the only way to maintain Freeze's analysis would be to propose a covert P°. In this vein, Kempchinsky (1995) proposes that (25) has the underlying structure in (26).

- (25) There are two cats in the tree.  
 (26)



For Kempchinsky, all existential structures have a covert P°. As in Freeze's proposal, this P° selects for a theme argument (*two cats*) and a location argument (*there in the tree*).

### 1.3.2 What about *haver*?

If Kempchinsky's proposal is on the right track, I have to assume that existential sentences with *haver* are derived in the same way, with *haver* being the spell-out of the copula plus an incorporated P°. This raises a question: why is the spell-out sometimes *ter* and sometimes *haver*? A more troublesome question seems to be why *ter* is the only verb available for *have* predications<sup>25</sup>. Kempchinsky's way out is proposing two different P°, an 'existential' P° and a 'possessive' P°. The incorporation of an existential P<sub>exist</sub> into a copula would spell out as *ter* or *haver*; but the incorporation of a possessive P<sub>poss</sub> can only spell out as *ter*.

A structural change that's happening in BP existentials might provide some evidence of the existence of P<sub>exist</sub> and P<sub>poss</sub>. Duarte (2003) reports an increasing tendency for the preverbal position of Brazilian Portuguese existentials to be filled. She suggests this is an effect of an ongoing change in the Null Subject Parameter, which is changing into a full-subject language, including a full phonological realization of expletive subjects. The use of dislocated elements (e.g. topicalized DPs, adverbs, PPs) to the left of the verb is common to both *ter* and *haver*. However, a strategy that is restricted to *ter* existentials is filling the empty preverbal position with a pronoun, mainly *você* ("you"), as in (27), or a locative or temporal DP, as in (29)<sup>26</sup>. (28) and (30) show that constructions with existential *haver* can be used as normal alternatives for (27) and (29) respectively.

- (27) Você tem um índice de audiência no horário eleitoral gratuito imenso.  
You have.PRES.3SG a rate of audience in.the time electoral free immense  
"There's an immense rating during the free electoral time."  
Literally: "You have an immense rating during the free electoral time."

- (28) Há um índice de audiência no horário eleitoral gratuito imenso.  
have.PRES.3SG a rate of audience in.the time electoral free immense.  
"There's an immense rating during the free electoral time."

---

<sup>25</sup> Contrast (19) with the ungrammaticality of the following sentence:

i) \*Camila há uma casa em Campos.  
Camila have.PRES.3SG a.FEM house in Campos  
"Camila has a house in Campos."

<sup>26</sup> Examples #27 and 29 were taken from the CETENFolha corpus.

(29) Março teve 493 apartamentos novos à venda na região.  
 March have.PAST.3SG 493 apartment.pl new.PL to.the.FEM sale in.the.FEM region.  
 "In March there were 493 new apartments for sale in the region."  
 Literally: "March had 493 new apartments for sale in the region."

(30) Em março houve 493 apartamentos novos à venda na região.  
 In March have.PAST.3SG 493 apartment.PL new.PL to.the.FEM sale in.the.FEM region.  
 "In March there were 493 new apartments for sale in the region."

In (27), *você* is a generic (or possibly non-referential) pronoun which occupies the Spec-IP position instead of the more common null expletive pronoun (as in (28)). Following Kempchinsky's analysis, changed only slightly to allow P° to take a temporal DP as argument, (29) would be the result of the lower P' raising to Spec-IP and P incorporating into the copula<sup>27</sup>. It should be noted that, although *ter* could be used in (28) and (30), *haver* cannot be used in (27) or (29). This might be evidence that (28) and (30) contain a P<sub>exist</sub>, which can be spelled out either as *ter* or *haver*; however, (27) and (29) contain a P<sub>poss</sub>, which can only be spelled out as *ter*.

The existence of two P° would also explain the fact that (31)<sup>28</sup> is ambiguous but (32) can only have the existential reading.

(31) Sempre tem um jogador no gancho, outro machucado.  
 always have.PRES.3SG a player in.the hook another injured  
 "There's always a player in the hook, another injured."  
 or: "(Corinthians) always has a player in the hook, another injured."

(32) Sempre há um jogador no gancho, outro machucado.  
 always have.PRES.3SG a player in.the hook another injured  
 "There's always a player in the hook, another injured."

Since the incorporated P<sub>exist</sub> can be spelled as either *ter* or *haver*, Kempchinsky's analysis seems to depend on both existential verbs occurring in free variation. Section 5.3 will show evidence that challenges this assumption; therefore, it seems that the adequacy of Freeze's hypothesis (as modified by Kempchinsky) is still an open question.

<sup>27</sup> Of course, now it's necessary to explain what happened with the P° *em* that has mysteriously disappeared (given that it's the covert P° that incorporated into the copula).

<sup>28</sup> Example taken from the CETENFolha corpus.

## 2. The definiteness effect

Since Milsark (1974), it has been noted that definite DPs appear to be barred from the pivot position of English existential sentences. Since then, several accounts have tried to explain what it is about definite DPs that prevents their appearance in that position. As mentioned in the previous chapter, this effect has also been documented in other constructions (e.g. post-posed subjects of unaccusatives) and across languages, including BP.

### 2.1 Strong vs. weak determiners

Milsark (1974) divided the class of DP types in two groups, according to which are permissible and which are barred in the pivot of an existential. He termed those groups "weak" and "strong" respectively<sup>29</sup>:

WEAK	STRONG
<i>a</i>	"definites"
<i>sm</i> <sup>30</sup>	<i>the</i>
number determiners	demonstratives
∅ plural and mass determiner in nonuniversal reading	pronouns
	possessive DET's
	universals
	<i>all</i>
	<i>every</i>
	<i>each</i>
	<i>any</i> when not polarity item of <i>some</i>
	∅ DET in universal reading

---

<sup>29</sup> Table (13) from Milsark (1977), p. 8.

<sup>30</sup> Milsark uses a distinction which had been noticed in Postal (1966). In sentence (i), *some* may have two meanings. In the first one, the act of walking in "was performed by some indeterminate but probably not large number of salesmen" (Milsark uses the notation *sm* to refer to this meaning of *some*). In the second meaning, "some subset of appropriate size to be referred to as 'some' has performed the action of entering, and [the sentence] carries a strong suggestion that some other group, by contrast, remained outside or is in some other way excluded from the situation described. The meaning of *some* in such cases might be paraphrased 'some (but not others)'" (p. 18).

(i) Some salesmen walked in.

Milsark then extends this distinction for other determiners, such as *many*, which also has a strong and a weak sense (for which he uses the notation *mny*).

According to Milsark, the property that all strong determiners share is that they are expressions of quantification. A strong determiner is an operator over the set of entities that picks out the members of that set for which the predication holds. On the other hand, all weak determiners can be analyzed as expressions of cardinality. A weak determiner expresses the size of the set of entities denoted by the DP.

Milsark assumes that the expression *there be* in an existential is an expression of existential quantification. Existentials with strong determiners are ungrammatical because there are two quantifications on the DP, one from the existential and another from the strong determiner. In contrast, existentials with weak determiners do not clash because there is only one quantification on the DP, the one from the existential.

## 2.2 A non-categorical effect

Most proposals have assumed that existential sentences constructed with a definite NP are ungrammatical. However, there are cases in which the definiteness effect can be overridden. For example, according to Viotti (1999), in a sample of existential sentences from the Projeto NURC spoken corpus, only 78% of them (as opposed to 100%) had an indefinite pivot. Milsark himself presented the following example in a footnote (even though he admitted that he didn't have anything very defensible to say about these cases):

(33) What else is there to worry about? Well, there's the wolf at the door...

Given the non-categoriality of the definiteness effect, Mikkelsen (2002) proposes an account within the framework of Optimality Theory. In this account, the definiteness effect is a consequence of constraints governing the subject position. Mikkelsen uses the following prominence scales (from Aissen 1999):

(34) Relational scale: Subject > Non-subject

(35) Definiteness scale: Definite > Strong Indefinite > Weak Indefinite

These scales formalize the intuition that subjects are more prominent than non-subjects, and that definite DPs are more prominent than indefinite ones. From the alignment of these two scales, the following harmony scales are obtained:

(36) Su/Def > Su/SI > Su/WI

(37) Non-su/WI > Non-su/SI > Non-su/Def

These harmony scales are mirrored in the following constraint hierarchies, which have the effect of penalizing the less harmonic alignments:

(38) \*Su/WI » \*Su/SI » \*Su/Def

(39) \*Non-su/Def » \*Non-su/SI » \*Non-su/WI

The harmony scale in (36) and the constraint hierarchy in (38) reflect the universal preference for definite subjects; on the other hand, indefinite DPs appearing in subject position will be dispreferred. Existentials are among the strategies that a language has to prevent the violation of the higher-ranked constraints in (38) and, consequently, existentials will have a preference for indefinite pivots. The mirror effect (and exceptions to it) should also be expected, namely, that canonical sentences will exhibit a strong preference for definite subjects.

Nevertheless, any given constraint can be overridden if a higher-ranked constraint of the grammar is at play. The exceptions to general tendencies (such as the definiteness effect) are cases in which the violation of a lesser constraint is preferable due to some higher-ranked constraint taking precedence.

### 3. Methodology

Mikkelsen (2002)'s account focused on Danish, leaving the cross-linguistic applicability of her proposal for future research. Since then, the Existential Research Group (a joint Stanford/Berkeley group research project) has formed, using corpus data to establish the distributional patterns of different DP-types in both existentials and canonicals across languages. As part of this research project, I performed a quantitative study on the CETENFolha corpus in order to establish the frequency and distribution of DP-types in both constructions in BP.

#### 3.1 The CETENFolha corpus

The CETENFolha corpus is a collection of texts from *A Folha de São Paulo*, one of Brazil's major newspapers, with a total of about 24 million words. The texts are selections of the articles published during 1994. The corpus is available in two versions, one of which is tagged for part of speech.

### 3.2 The search templates

I used three basic templates<sup>31</sup>, one for canonicals (40) and two for existentials (41 and 42).

(40) DP + *estar* + *em*

(41) *haver* + DP + *em*

(42) *ter* + DP + *em*

**DP-types:** I searched for 21 DP-types, 16 of them classified according to the head determiner and 5 lexical DP categories (personal pronouns and other pronominalized DPs). The complete list of DP-types is found in (43).

(43) Indefinite DPs:

*um* ("a")

*algum* ("some")

*nenhum* ("no")

*muito* ("much" or "many")

*pouco* ("few")

Numerals

Definite DPs:

*o* ("the")

*este* ("this")

*esse* ("that")

*aquele* ("that" – distal)

*todo* ("every")

*todos os* ("all the")

*a maioria dos* ("most of the")

Numerals + *dos* ("of the")

*algum dos* ("some of the")

*muitos dos* ("many of the")

Indefinite lexical DPs:

*algo/algúém* ("something"/"somebody")

*nada/ninguém* ("nothing"/"nobody")

Definite Lexical DPs:

Personal Pronouns

*tudo* ("everything")

*isto/isso/aquilo* ("this one"/"that one"/"that one" – distal)

**The internal structure of the DP:** Using regular expressions syntax, I was able to capture complex DPs containing optional PP modifiers as well as both pre- and post-posed AP

---

<sup>31</sup> The actual search expressions can be found in Appendix I.

modifiers. Besides, the use of the disjunction operator ( $\cup$ ) allowed me to abstract away from the different gender and number variants of the determiner.

**The verb:** I used lexemes to find all possible tense and mood variants of *haver* and *estar*. However, I preferred the use of the disjunction operator for *ter* in order to restrict the matches to the 3SG forms of the verb (the only forms found in impersonal existential constructions).

**The PP:** The use of the preposition *em* ("in") in the search templates is due to two main reasons: i) an attempt to restrict the number of false hits that would have resulted from a less restricted canonical search, avoiding sentences of the "copula + DP" or "copula + AP" type; and ii) an effort to obtain "minimal triplets," constructions that are as structurally similar as possible and allow –at least in theory– the existential/canonical alternation. Since *em* ("in") is the most widely used preposition, the restricted search still resulted in a sample big enough to reflect the distributional patterns of most NP-types.

## 4. Results

### 4.1 Preliminary numbers

The 63 searches returned a total of 9138 sentences. In spite of the search restrictions, this number still included many inappropriate matches. These were mainly due to the homophony between existential *ter* and possessive *ter*. Of the 2351 sentences with *ter*, only 116 were existential constructions<sup>32</sup>.

The second main source of inappropriate matches was due to DPs that immediately preceded the verb *estar* but were embedded within a bigger DP (e.g. the second half of a coordinated DP). For example, sentences like (44) and (45) were discarded because the structure of the search expressions would have precluded the match of existentials with such complex DPs as *a metralhadora e a espingarda* (coordination) or *a separação de Zamora com a primeira mulher* (two PPs) in pivot position. From the original 4923 canonical sentences, 951 were inappropriate matches of this sort.

---

<sup>32</sup> In fact, sentences (18) and (19) are actual examples picked out by the search expressions.

(44) A metralhadora e a espingarda estavam na sala.<sup>33</sup>  
the.FEM machine.gun and the.FEM shotgun be.PAST.3PL in.the lounge  
"The machine gun and the shotgun were in the lounge."

(45) A separação de Zamora com a primeira mulher estava em litígio.  
the.FEM separation of Zamora with the.FEM first.FEM woman be.PAST.3SG in dispute  
"Zamora's separation from [his] first wife was in dispute."

## 4.2 Adjusted numbers

After discarding the inappropriate matches, there remained a total of 5924 sentences. Additionally, the matches for 4 DPs (*aquele*, "that one" – distal; *muitos dos*, "many of the;" *todo*, "every;" and *alguns dos*, "some of the") were also discarded because their numbers were too low to have any statistical relevance (the number of matches was 2, 5, 4, and 5 respectively).

The remaining 5908 sentences had the following distribution:

- 3959 canonicals
- 1949 existentials: 1834 with *haver*  
115 with *ter*

Such low occurrence of *ter* in existential sentences is surprising, even under the assumption that *haver* is the preferred verb for written BP. Moreover, the majority of occurrences with *ter* are examples of direct quotations. Further research comparing the CETENFolha corpus with another written corpus from a different genre would be required to find out if these numbers reflect an actual tendency of written BP instead of just an editorial policy of *A Folha de São Paulo*. In any case, it seems reasonable to claim that written BP lags far behind in regards to the language changes that have been documented for spoken BP.

## 4.3 Existential ratios

In order to study the relative frequency of any DP-type in canonical and existential constructions, I calculated the ratio of the number of canonical sentences to the number of existential sentences for each DP-type. To exemplify how the ratios are calculated, let's consider two examples.

---

<sup>33</sup> All examples in this chapter were taken from the CETENFolha Corpus.

After running the three searches with the Determiner *um* ("a") and discarding the inappropriate matches, there were 1107 hits for the two existential constructions, and there were 48 hits for the canonical construction. I then calculated the ratio of the adjusted canonical frequency to the adjusted existential frequency (48 divided by 1107), which gave me an existential ratio of 0.04. The existential occurrences of DPs headed by *um* outnumber the canonical occurrences by a factor of more than 20, and the interpretation of the 0.04 ratio is that *um* is strongly existential.

In the case of the Determiner *o* ("the"), after discarding the inappropriate matches from the three searches, there were 84 hits for the two existential constructions and 2861 hits for the canonical construction. The ratio of the adjusted canonical frequency to the adjusted existential frequency (2861 divided by 84) is 34.06. The canonical occurrences of DPs headed by *o* outnumber the existential occurrences by a factor of more than 30, which means that *o* is strongly canonical.

For those DPs that had no occurrences in pivot position, I treated the ratio as tending to infinity (represented by  $\infty$ ).

The table on the following page shows the number of occurrences for each DP as well as their existential ratios<sup>34</sup>.

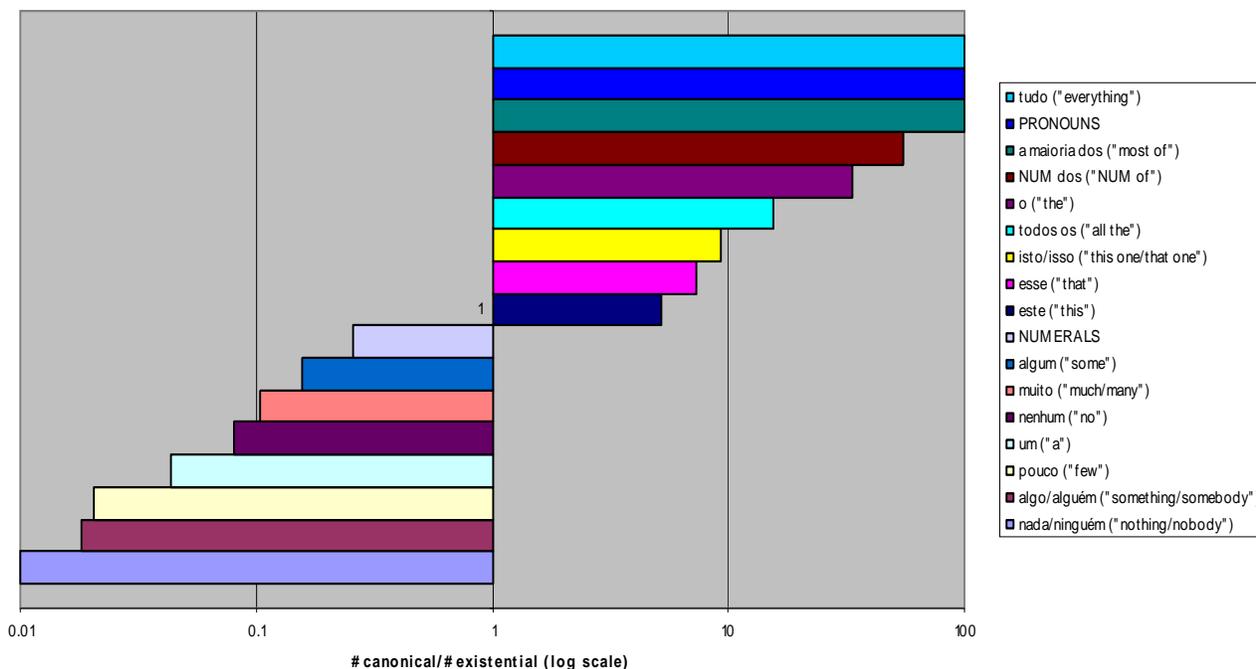
---

<sup>34</sup> A table comparing the Preliminary # of hits to the Actual # of hits can be found in Appendix II.

DP-type	Actual # of hits			Total	Existential ratio
	Existentials		Canonicals		
	<i>haver</i>	<i>ter</i>	<i>estar</i>		
nada/ninguém ("nothing"/"nobody")	30	11	0	41	0.00
algo/alguém ("something"/"somebody")	45	10	1	56	0.02
pouco ("few")	48	1	1	50	0.02
um ("a")	1057	50	48	1155	0.04
nenhum ("no")	49	1	4	54	0.08
muito ("much"/"many")	116	19	14	149	0.10
algum ("some")	68	8	12	88	0.16
NUMERALS	324	7	85	416	0.26
este ("this")	6	0	31	37	5.17
esse ("that")	8	1	66	75	7.33
isto/isso ("this one"/"that one")	3	0	27	30	9.00
todos os ("all the")	2	0	31	33	15.50
o ("the")	78	6	2861	2945	34.06
NUM dos ("NUM of the")	0	1	55	56	55.00
a maioria dos ("most of the")	0	0	18	18	∞
PRONOUNS	0	0	686	686	∞
tudo ("everything")	0	0	19	19	∞
	1834	115	3959	5908	

**TABLE 3: Number of occurrences of each DP-type in existentials and canonicals.**

The graph in the following page presents the existential ratios for each DP in a clearer way (the bars and the legend follow the same vertical order):



**GRAPH 1: Existential ratios per DP-type.**

#### 4.4 Definiteness effect

The graph above shows a strong definiteness effect. The left side contains those DP-types that appear more frequently in the pivot position of an existential. All of them are indefinite or, in Milsark's terminology, weak DPs.

It's important to remember that the chart above shows relative frequency (frequency of occurrence in subject position divided by frequency of occurrence in pivot position), which is a more reliable indication of the existentiality of a DP than absolute frequency. However, the absolute percentages also show clear evidence of the definiteness effect. Of 1949 existential sentences, 95% had an indefinite DP in pivot position. 57% of existentials had a DP headed by *um* ("a") and another 17% were headed by a numeral.

Nevertheless, it can also be observed that the definiteness effect is non-categorical. In spite of the strong preference for pivots to be indefinite DPs, there were 105 cases of definite pivots, such as the example in (46):

(46) Em relação ao ano passado, quando houve **a inflexão à esquerda** no partido...  
 in relation to.the year last when have.PAST.3SG **the.FEM turning to.the.FEM left** in.the party

"Compared to last year, when there was the change to the left in the party..."

Moreover, as (47) to (49) show, the three DPs that appear to be categorically canonical (*a maioria dos*, "most of the;" pronouns; and *tudo*, "everything") stop being so if the requirement of a PP coda is dropped.

(47) Lá há **a maioria das casas de striptease e prostituição**.  
 there have.PRES.3SG **the majority of.the.FEM.PL house.PL of striptease and prostitution**

"There are most of the striptease and prostitution houses there."

(48) Houve **tudo**, só não houve crime.  
 have.PAST.3SG **everything** only NEG have.PRES.3SG crime

"There was everything, only there wasn't crime."

(49) Entre os novos alunos, tinha **eu** e alguns alemães.  
 between the.PL new.PL pupil.PL, have.PAST.3SG **I** and some.PL German.PL

"Among the new pupils, there was me and some Germans."

#### 4.5 Indefiniteness effect

The data also confirms that canonical sentences show the INDEFINITENESS EFFECT, a strong preference to have a definite DP in subject position. As can be seen on the right-side of Graph 1, the DPs that appear more frequently as subjects of canonicals are the definite, or strong, DPs.

In absolute terms, 96% of the 3959 canonical sentences had a definite DP in subject position. The most common subjects were DPs headed by *o* ("the," 72%) and personal pronouns (17%).

The indefiniteness effect is also a non-categorical effect. 165 cases were found of indefinite DPs in subject position, such as the example in (50):

- (50) **Um sujeito** está na feira vendendo vasos.  
**a person** be.PRES.3SG in.the.FEM fair sell.GER pot.PL  
 "A person is in the fair selling pots."

## 5. Towards an OT analysis

The results in Chapter 4 confirm that the definiteness and indefiniteness effects exist in BP, but also that they are not categorical. This latter fact provide evidence in favor of an optimality-theoretic approach such as the one in Mikkelsen (2002). Since Mikkelsen only considers three broad DP-types (definites, strong indefinites, and weak indefinites), she is able to account for the definiteness effect in Danish using the harmony scale in (51) and the constraint hierarchy in (52) (repeated from (36) and (38) above).

- (51) Su/Def > Su/SI > Su/WI  
 (52) \*Su/WI » \*Su/SI » \*Su/Def

The hierarchies in (51) and (52) are not enough to account for the finer distinctions found in the corpus data.

For example, the complete distribution of DP-types represented in Graph 1 (and repeated in (53) in the form of an existential scale) shows that personal pronouns are more canonical than DPs headed by the definite article *o* ("the"), which are in turn more canonical than DPs headed by demonstratives such as *este* ("this") and *esse* ("that"). These three kinds of DPs fall under Mikkelsen's class of Definite DPs so there is no way within her approach to distinguish them.

- (53) *tudo* ("everything") > **PERSONAL PRONOUNS** > *a maioria* ("most") > ***o* ("the")** > *todos* ("all") > ***isto/isso* ("this one"/"that one")** > *esse* ("that") > *este* ("this") > NUMERALS > *algum* ("some") > *muito* ("much"/"many") > *nenhum* ("no") > *um* ("a") > *pouco* ("few") > *algo/algúem* ("something"/"somebody") > *nada/ninguém* ("nothing"/"nobody")

A first step to provide a full account of the distribution in (53) is the use of an extended definiteness scale, such as the one in (54), proposed in Aissen (2003).

(54) Definiteness scale: Personal pronoun > Proper name > Definite DP > Indefinite Specific DP > Non-specific DP

(55) Relational scale: Subject > Non-subject

Through the alignment of the scale in (54) with the Relational scale in (55) (repeated from (34) above), the harmony scales in (56) and (57) and the constraint hierarchies in (58) and (59) are obtained.

(56) Su/PersPron > Su/ProperName > Su/Def > Su/IndefSpec > Su/IndefNon-spec

(57) Non-su/IndefNon-spec > Non-su/IndefSpec > Non-su/Def > Non-su/ProperName > Non-su/PersPron

(58) \*Su/IndefNon-spec » \*Su/IndefSpec » \*Su/Def » \*Su/ProperName » \*Su/PersPron

(59) \*Non-su/PersPron » \*Non-su/ProperName » \*Non-su/Def » \*Non-su/IndefSpec » \*Non-su/IndefNon-spec

The use of the extended Definiteness scale is only a first step towards providing a full account of the distribution of DPs in BP. The personal pronouns and the proper names (a category I didn't investigate in this corpus study) have been separated from the Definite DP category. However, there is still the need to create a category for the demonstratives. Is there enough cross-linguistic evidence as to where in the hierarchy they should be? The BP data as well as the English data reported in Beaver et al. (to appear) suggest that demonstratives should be below definite DPs in the hierarchy. On the other hand, the Dutch data (also from Beaver et al.) suggests the opposite. Of course, it is also necessary to consider evidence from other phenomena (and languages) in order to establish a universal hierarchy.

A related issue is: how many categories should be created for the demonstratives? If we take differences in distribution (such as the existential ratio) as a criterion for proposing new categories, what is a sufficient difference to merit the creation of a category? Is it possible that we lose greater generalizations by proposing such fine-grained hierarchies? For example, do all demonstratives belong to a natural class? In the case of BP, *este* ("this") and *esse* ("that") have sufficiently similar distributions that it's reasonable to group them in one single category. However, it's not possible to have a complete picture of BP demonstratives based on this corpus study since there wasn't enough data to analyze the distribution of *aquele* ("that" – distal). Should this be taken as evidence in itself that they don't form a natural class? How does

frequency of use determine syntactic parallelism between constructions? These questions are, unfortunately, outside the scope of this thesis and await further investigation.

Definiteness is not the only relevant property to decide what makes a good subject. As seen in Chapter 4, both definite and indefinite DPs can appear in the subject position of a canonical and the pivot position of an existential. A closer look at the exceptions suggests that the Animacy scale (60) might also be relevant in the existential/canonical alternation.

(60) Animacy scale:      Human > Animate non-human > Inanimate

As Table 4 shows, from a sample of 84 definite pivots, only 1% had a human referent – (61) is the single example of a human definite pivot; 99% of them had an inanimate referent (as in (62), repeated from (46) above). The proportions are quite different for definite subjects: 22% had a human referent and 78% had an inanimate referent<sup>35</sup>.

It should also be noted that (61) does not have a singular definite reference, but rather seems to have a generic reference. It is possible that this relates to a characteristic of definite DPs in Romance languages, which have a broader use than in English, as suggested to me by Richard Rhodes.

DPs headed by <i>o</i> , "the"	Total	Human		Inanimate	
		Number	Percentage	Number	Percentage
Canonicals	2878	621	22%	2242	78%
Existentials	84	1	1%	83	99%

**TABLE 4: Distribution of definite DPs according to animacy.**

<sup>35</sup> There are 16 examples of definite DPs with an animate non-human referent, all of them in subject position. Given their low number, I didn't include a column for them.

(61) Não se pode falar em avanços se não houver **a mulher** em todos  
 NEG REFL can.3SG speak in advance.PL if NEG have.INF **the.FEM woman** in every.PL  
 os níveis de participação e, principalmente, de decisão.  
 the level.PL of participation and mainly of decision  
 "It's not possible to speak of advances unless women are in every level of participation  
 and, mainly, of decision."  
 Literally: "unless there is the woman in every level..."

(62) Em relação ao ano passado, quando houve **a inflexão à**  
 in relation to.the year last when have.PAST.3SG **the.FEM turning to.the.FEM**  
**esquerda** no partido...  
**left** in.the party  
 "Compared to last year, when there was the change to the left in the party..."

Since the corpus is a collection of texts from a newspaper, the preponderance of inanimate referents (mainly abstract) doesn't come as a surprise. However, for a definite DP with a human referent, the occurrences in subject position outnumber the occurrences in pivot position by a factor of 621. This number should be contrasted with the distribution of definite DPs with an inanimate referent: given the large number of canonicals, the occurrences in subject position still outnumber the occurrences in pivot position, but only by a factor of less than 30. A DP with an inanimate referent is relatively more common in pivot position than in subject position, which suggests that animacy is indeed a factor in the existential/canonical alternation.

If the Animacy scale were relevant to the existential/canonical alternation, we would further expect that indefinite subjects with a human referent would be relatively more frequent than indefinite pivots with a human referent. This is indeed the case. As Table 5 shows, from a sample of 48 indefinite subjects, 35% of them had a human referent<sup>36</sup>; as for the indefinite pivots, only 5% of them had a human referent. In other words, there's a preference for a subject to have a human referent and this preference might override the preference for a definite subject.

---

<sup>36</sup> There's only one example of an indefinite subject with an animate non-human referent. As for Table 4, given the low number, I didn't include a column for Animate non-human.

DPs headed by <i>um</i> , "a"	Total	Human		Inanimate	
		Number	Percentage	Number	Percentage
Canonicals	48	17	35%	30	63%
Existentials	1107	57	5%	1050	95%

**TABLE 5: Distribution of indefinite DPs according to animacy.**

The Animacy scale seems also to be relevant for the choice of existential verb. As Table 6 shows, human referents are relatively more frequent with *ter* than with *haver*.

DPs headed by <i>um</i> , "a"	Total	Inanimate		Human	
		Number	Percentage	Number	Percentage
Existentials with <i>ter</i>	50	39	78%	11	22%
Existentials with <i>haver</i>	1057	1011	96%	46	4%

**TABLE 6: Distribution of inanimate and human referents in existentials.**

Even though these are clearly tendencies and not absolute effects, they seem to suggest that *ter* and *haver* are not really in free variation; *ter* is preferred for pivots with human referents and *haver* for inanimate referents. However, there is another issue that might be relevant for the distribution in Table 6. Most of the sentences with *ter* are cases of direct quotations; therefore, the distribution above could be a reflection of the preponderance of human referents in the particular discourses that were quoted or in spoken language in general.

## 6. Conclusion

This thesis has investigated the existential/canonical alternation in BP. Chapter 1 argued against the analysis of existentials as unaccusatives and in favor of considering them part of the impersonal constructions. It also presented Freeze's hypothesis that existentials are part of the locative paradigm, and it suggested that this hypothesis could be reconciled with the data at the cost of proposing more abstract structure. Chapter 2 presented different accounts of the definiteness effect and pointed out the problems most of these face: the non-categoricity of the

effect. It also presented Mikkelsen's alternative analysis within Optimality Theory, which allows for exceptions to be accounted for. This thesis has attempted to investigate if her analysis could be extended to BP; in order to do that, it was necessary to establish the distribution of several DP-types in BP existentials and canonicals. Chapters 3 and 4 presented the methodology and results from a quantitative study performed on the CETENFolha corpus of written BP data, which confirmed the existence of the definiteness effect in BP existentials. There's also evidence of the mirror effect, an indefiniteness effect in BP canonicals. However, neither of these effects is categorical since all DP-types occur in both constructions. Chapter 5 suggested that an OT analysis is the more appropriate approach to provide a full account of these effects and its exceptions. It also presented evidence that both the Definiteness and the Animacy scales are relevant for the existential/canonical alternation.

Further investigation is needed on the relationship between existential *ter* and possessive *ter* as well as the factors determining the choice of existential verb. These issues seem to be the most relevant to examine the adequacy of a transformational analysis à la Freeze (or the need for a lexicalist account).

It is also necessary to investigate the precise interaction between the constraints that are obtained from the Definiteness and Prominence scales on one side, and the Animacy and Prominence scales on the other. Another open question is how other subject properties, such as topicality, affect the canonical/existential alternation. These issues await further research.

## References

- Aissen, Judith (2003). "Differential Object Marking: Iconicity vs. Economy." *Natural Language and Linguistic Theory* 21: 435-483.
- Beaver, David, Itamar Francez and Dmitry Levinson (to appear). In E. Georgala and J. Howell (eds.), *Proceedings of Semantics and Linguistic Theory XV*, CLC Publications, Ithaca, New York.
- Callou, Dinah and Juan O. Avelar (2001). "Sobre ter e haver em construções existenciais: variação e mudança no português do Brasil." *Revista Gragoatá* 9: 85-100.
- Duarte, Maria Eugênia (2003). "O sujeito expletivo e as construções existenciais." In: Roncarati, Cláudia & Abraçado, Jussara, eds., *Português brasileiro – contato lingüístico, heterogeneidade e história*, 123-131. Rio de Janeiro: 7 Letras/Faperj.
- Franchi, C.; Negrão, E. V.; Viotti, E. (1998). "Sobre a Gramática das Orações Impessoais com *Ter/Haver*." *Delta* 14: 105-131.
- Freeze, Ray (1992). "Existentials and Other Locatives." *Language* 68: 553-595.
- Guéron, Jacqueline (1980). "On the syntax and semantics of PP extraposition." *Linguistic Inquiry* 11: 637-678.
- Kempchinsky, Paula (1995). "Perfective Auxiliaries, Possession and Existence in Romance." In: Zagona, Karen, ed., *Grammatical Theory and Romance Languages*, 135-144. Amsterdam: John Benjamins Publishing Company.
- Mikkelsen, Line (2002). "Reanalyzing the definiteness effect: evidence from Danish." *Working Papers in Scandinavian Syntax* 69: 1-75.
- Milsark, Gary L. (1974). *Existential sentences in English*. Doctoral dissertation. Cambridge: MIT.
- Milsark, Gary L. (1977). "Toward an Explanation of Certain Peculiarities of the Existential Construction in English." *Linguistic Analysis* 3: 1-29.
- Nascimento, Milton do (1984). *Sur la posposition du sujet dans le Portugais du Brésil*. Doctoral dissertation. Paris: Université de Paris VIII.

- Perlmutter, David (1976). "Evidence for Subject Downgrading in Portuguese." In: J. Schmidt-Ratefeldt, ed., *Reading in Portuguese Linguistics*, 93-138. Amsterdam: North-Holland.
- Postal, Paul (1966). "On So-called 'Pronouns' in English. " In Francis P. Dinneen, ed., *Problems in Semantics, History of Linguistics, Linguistics and English*, 178-206. Washington, DC: Georgetown University Press.
- Silva, María Cristina Figueiredo (1994). *La position sujet en Portugais Brésilien*. Doctoral dissertation. Génève: Université de Génève.
- Suñer, Margarita (1982). *Syntax and Semantics of Spanish Presentational Sentence-Types*. Washington: Georgetown University Press.
- Thomas, Earl W. (1969). *The syntax of spoken Brazilian Portuguese*. Nashville: Vanderbilt Univ. Press.
- Viotti, Evani (1999). *A sintaxe das sentenças existenciais do português do Brasil*. PhD Thesis, Departamento de Lingüística, Universidade de São Paulo.
- Viotti, Evani (2002). "A Estrutura Sintática das Sentenças Existenciais e o Efeito de Definitude: Semelhanças e Diferenças entre o Inglês e o Português do Brasil." *Revista Letras* 58: 371-395.
- Zubizarreta, María Luis (1982). "Theoretical implications of subject extractions in Portuguese." *The Linguistic Review* 2: 79-96.

## Appendix I: Search templates

### Canonical template (with *D o*, "the"):

```
"[OoAa]s?" [pos="ADV.*"]* [pos="ADJ.*"]* [pos="ADV.*"]* [pos="N.*"]+ [pos="ADV.*"]*  
[pos="ADJ.*"]* [pos="ADV.*"]* [pos="PRP.*"]* [pos="*.DET.*"]* [pos="ADV.*"]*  
[pos="ADJ.*"]* [pos="ADV.*"]* [pos="N.*"]* [pos="ADV.*"]* [pos="ADJ.*"]*  
[pos="ADV.*"]* [pos="V.*"]* [lema="estar"] "em|n[oa]s?|nu[mn]s?|numas?" [classe="JOCF"]37
```

### Existential template for verb *haver* (with *D o*, "the"):

```
[lema="haver"] "[oa]s?" [pos="ADV.*"]* [pos="ADJ.*"]* [pos="ADV.*"]* [pos="N.*"]+  
[pos="ADV.*"]* [pos="ADJ.*"]* [pos="ADV.*"]* [pos="PRP.*"]* [pos="*.DET.*"]*  
[pos="ADV.*"]* [pos="ADJ.*"]* [pos="ADV.*"]* [pos="N.*"]* [pos="ADV.*"]*  
[pos="ADJ.*"]* [pos="ADV.*"]* "em|n[oa]s?|nu[mn]s?|numas?" [classe="JOCF"]
```

### Existential template for verb *ter* (with *D o*, "the"):

```
"[Tt]e[mr]||[Tt]eve|[Tt][ie]nha|[Tt]ivera?|[Tt]erá|[Tt]eria|[Tt]ivesse|[Tt]endo|[Tt]ido" "[oa]s?"  
[pos="ADV.*"]* [pos="ADJ.*"]* [pos="ADV.*"]* [pos="N.*"]+ [pos="ADV.*"]*  
[pos="ADJ.*"]* [pos="ADV.*"]* [pos="PRP.*"]* [pos="*.DET.*"]* [pos="ADV.*"]*  
[pos="ADJ.*"]* [pos="ADV.*"]* [pos="N.*"]* [pos="ADV.*"]* [pos="ADJ.*"]*  
[pos="ADV.*"]* "em|n[oa]s?|nu[mn]s?|numas?" [classe="JOCF"]
```

---

<sup>37</sup> [classe="JOCF"] is used to restrict the search to the CETENFolha corpus.

## Appendix II: Number of hits per DP-type

	Preliminary # of hits			Actual # of hits			To
	Existentials		Canonicals	Existentials		Canonicals	
	haver	ter	estar	haver	ter	estar	
nada/ninguém (" <i>nothing/nobody</i> ")	31	22	3	30	11	0	
algo/alguém (" <i>something/somebody</i> ")	48	20	2	45	10	1	
pouco (" <i>few</i> ")	48	33	6	48	1	1	
um (" <i>a</i> ")	1076	1064	120	1057	50	48	
nenhum (" <i>no</i> ")	49	38	5	49	1	4	
muito (" <i>much/many</i> ")	116	74	19	116	19	14	
algum (" <i>some</i> ")	70	57	14	68	8	12	
NUMERALS	324	611	417	324	7	85	
este (" <i>this</i> ")	6	8	44	6	0	31	
esse (" <i>that</i> ")	8	17	80	8	1	66	
isto/isso (" <i>this one/that one</i> ")	3	3	44	3	0	27	
todos os (" <i>all the</i> ")	2	17	46	2	0	31	
o (" <i>the</i> ")	79	359	3286	78	6	2861	
NUM dos (" <i>NUM of</i> ")	0	12	48	0	1	55	
a maioria dos (" <i>most of</i> ")	0	2	18	0	0	18	
PRONOUNS	0	3	732	0	0	686	
tudo (" <i>everything</i> ")	0	4	23	0	0	19	
	1860	2344	4907	1834	115	3959	